Office de la propriété intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An Agency of Industry Canada

#9
D. J.
10-18-0

Bureau canadien des brevets
Certification

La présente atteste que les documents ci-joints, dont la liste figure ci-dessous, sont des copies authentiques des documents déposés au Bureau des brevets.

Canadian Patent Office
Certification

This is to certify that the documents attached hereto and identified below are true copies of the documents on file in the Patent Office.

Specification and Drawings, as originally filed, with Application for Patent Serial No: **2,230,188**, on March 27, 1998, by **WILLIAM C. TREURNIET, LOUIS THIBAULT, SEYMOUR SHLIEN AND GILBERT ARTHUR JOSEPH SOULODRE**, for "Objective Audio Quality Measurement".

CERTIFIED COPY OF
PRIORITY DOCUMENT

Agent certificateur/Certifying Officer

June 30, 2004
Date

Canada

(CIPO 68)
04-09-02

OPIC CIPO

Objective Audio Quality Measurement

A reliable subjective quality assessment of audio or speech signals may be obtained from human listeners, but the process is labour-intensive and time-consuming. Listeners are typically asked to judge the quality of a processed audio or speech sequence relative to an original unprocessed version of the same sequence.

Current objective measures of audio or speech quality include THD (Total Harmonic Distortion) and SNR (Signal-to-Noise Ratio). The latter metric can be measured on either the time domain signal or a frequency domain representation of the signal. However, these measures are known to provide a very crude measure of audio or speech quality and are not well correlated with the subjective quality. This lack of correlation worsens when these metrics are used to measure the quality of devices such as A/D and D/A converters and perceptual audio (or speech) codecs which make use of the masking properties of the human auditory system. With these devices, audio (or speech) signals are often perceived to be of good or excellent quality even though the measured SNR may be poor.

An object of the present invention is to develop more efficient methodology for obtaining such quality ratings.

The invention utilizes a computer program which enables an objective assessment of the subjective quality of a processed audio sequence relative to an original unprocessed version of the same sequence. It assumes that both versions are simultaneously available in computer files and that they are synchronised in time. The audio sequences are processed by a computational model of hearing which removes auditory components from the input that are normally not perceptible by human listeners. The result is a numerical representation of the pattern of excitation produced by the sounds on the basilar membrane of the human auditory system. The basilar sensation level of the processed version is compared with that of the unprocessed version, and the difference is used to predict the average quality rating that would be expected from human listeners.

The advantage of this invention over other approaches to the problems noted above is that it provides a much more efficient means of obtaining an estimate of the perceptual quality of an audio or speech sequence that has been processed in some way. This permits frequent and automated monitoring of audio or speech equipment performance and the degree of communication network degradation. Although others have developed systems for measurement of objective perceptual quality of wide-band audio [8] [9] [10] [11], these employ algorithms that were shown to result in inadequate levels of performance in tests conducted by the ITU-R in 1995-6. The present invention constitutes an advance over these prior systems.

In its various aspects, the invention comprises the following novel concepts and methods:

- Improved middle ear attenuation spectrum to extend high frequency cutoff.

- Improved frequency-to-pitch mapping function.
- Improved spreading function slopes for improved performance.
- Recursive filter implementation of spreading function.
- Quantized, level-dependent spreading function using recursive filter.
- Frequency-dependent spreading function using recursive filter.
- Concept of separate weightings for adjacent frequency ranges.
- Concept of perceptual inertia.
- Establishing perceptual asymmetry.
- Concept of adaptive threshold for rejection of relatively low values.
- Concept of using Coefficient of Variation to differentially weight brief versus continuous distortions.
- Concept of reference and distortion spectra similarity affecting quality.
- Concept of harmonic structure in the distortion affecting quality.
- Concept of distortion activity affecting quality.
- Concept of time-frequency aggregation of distortion affecting quality.
- Measuring the temporal lag between channels as a function of frequency.
- Detection of pre-echo.
- Measuring stereo imaging.
- Use of a particular combination of variables for predicting audio quality.
- Concept of a neural network to generate a quality measure from model variables.

The invention can be used in a variety of applications, for example:

- Assessment of implementations - a procedure to characterize different implementations of audio processing equipment, in many cases audio codecs.

- Perceptual quality line up - a fast procedure for checking out a piece of equipment or a circuit before putting it into service.

- On-line monitoring - a continuous process to monitor an audio transmission in service.

- Codec development - a procedure for comparing competing encoding algorithms.

- Network planning - a procedure to optimize the cost and performance of a transmission network under given constraints.

- Aid to subjective assessment - a tool for screening critical material to include in a listening test.

In the accompanying drawings:

Fig. 1 is a high level representation of a computational model of audition developed as a tool for objective evaluation of the perceptual quality of audio signals; and

Fig. 2 shows successive stages of processing of the model.

A computational model of audition developed as a tool for objective evaluation of the perceptual quality of audio signals will be described. The description presents the component structure of the model, as well as details of the signal transformations performed by the components. Simulation of peripheral auditory processing results in a basilar membrane representation, and simple transformations, based on assumptions about higher level perceptual processing, lead to an estimated perceptual quality of the signal relative to a known reference signal. The model was calibrated using data obtained from human observers in a number of listening tests.

A high level representation of the model is shown in Fig. 1. Time domain versions of both a reference signal and the same signal possibly altered in some way (e.g., processed by a lossy compression algorithm) are transformed by the peripheral ear model to their respective representations on the basilar membrane. The basilar representation is called the *basilar sensation*. The *basilar degradation* is obtained by comparing the basilar sensation derived from the reference signal to that derived from the altered signal. The basilar degradation as a function of time is analysed by a *cognitive model* which outputs an objective perceptual quality rating based on the monaural degradations as well as any shifts in the position of the binaural auditory image.

The Auditory (Peripheral Ear) Model

The program models the transfer characteristics of the middle and inner ear to form an internal representation of the signal. The input signal is decomposed into a time-

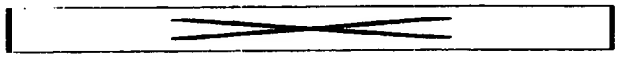
frequency representation using a discrete fournier transform (DFT). Typically, a Hann window of approximately 40 msec is applied to the input data, with a 50 percent overlap between successive windows. The energy spectrum is multiplied by a frequency dependent function which models the effect of the ear canal and the middle ear. The attenuated spectral energy values are mapped from the frequency scale to a pitch scale that is more linear with respect to both the physical properties of the inner ear and observed psychophysical effects. The transformed energy components are then convolved with a spreading function to simulate the dispersion of energy along the basilar membrane. Finally, an intrinsic frequency-dependent energy is added to each pitch component to account for the absolute threshold of hearing. Conversion of the energy to decibels results in a basilar membrane representation of the signal. These successive stages of processing are shown in Fig. 2.

The following computations are performed for each block:

- The energy spectrum is multiplied by the attenuation spectrum of a low pass filter which models the effect of the ear canal and the middle ear. The attenuation spectrum, described by the following equation, is modified from that presented in reference [3] in order to extend the high frequency cutoff. This was accomplished changing the exponent in the following formula from 4.0 to 3.6.

$$A_{dB} = -6.5\, e^{(-0.6(f-0.33)^2)} + 10^{-3} f^{3.6}$$

- The attenuated spectral energy values are transformed using a non-linear mapping function from the frequency domain to the subjective pitch domain using the bark scale (an equal interval pitch scale). A commonly used mapping function [5] is as follows.



A new function was created that improves resolution at higher frequencies. A simpler expression was derived for this new function, and is presented as follows (the frequency f is in Hz.):

$$p = f/(9.0304615e\text{-}05f + 2.6167612)$$

- The basilar membrane components are convolved with a spreading function to simulate the dispersion of energy along the membrane. The spreading function applied to a pure tone results in an asymmetric triangular excitation pattern with slopes that may be selected to optimize performance. The spreading is implemented by sequentially applying two IIR filters,

$$H_1(z) = 1/(1\text{-}a/z) \quad \text{and} \quad H_2(z) = 1/(1\text{-}bz)$$

where the a and b coefficients are the reciprocals of the slopes of the spreading function on the dB scale.

The original program implemented a spreading function with a slope on the low frequency side (LSlope) of 27 dB/Bark and a slope on the high frequency side of -10 dB/Bark. For the frequency-to-pitch mapping function given above, it has been found

that predictions of audio quality ratings improved with fixed spreading function slopes of 24 and -4 dB/Bark, respectively.

The literature (e.g. [3]) indicates that the slope S of the spreading function on the low frequency side should be fixed (i.e., *Lslope* = 0.27 dB/mel). However, on the high frequency side, the slope is said to vary with both signal level and frequency. That is, it increases with both decreasing level and increasing frequency. The following equation for computing the higher frequency slope (S) as a function of frequency and level is based on an equation in [3]:

$$S_{dB/mel} = HSlope - LRate \cdot L_{dB} + FRate/F_{KHz}$$

Suggested values of 24 for *HSlope*, 230 for *FRate* and 0.2 for L*Rate* . However, in the model, the best values for these parameters are dependent on other system components such as the frequency to pitch mapping function. The inventors find parameter values for a particular system configuration using a function optimization procedure based on a genetic algorithm. Optimal values are those that minimize the difference between the model's performance and a human listener's performance in a signal detection experiment. This procedure allows the model parameters to be tailored so that it behaves like a particular listener - reference [6].

In other psychoacoustic models, the spreading function is applied to each pitch position by distributing the energy to adjacent positions according to the magnitude of the spreading function at those positions. Then the respective contributions at each position are added to obtain the total energy at that position. Dependence of the spreading function slope on level and frequency is accommodated by dynamically selecting the slope that is appropriate for the instantaneous level and frequency.

To implement the dependence of the slope on level using the IIR filter implementation, a new procedure was developed. Note that, because convolution is a linear operation, the effects of convolving data with different spreading functions may be summed. Therefore, input values within particular ranges are convolved with level-specific spreading functions, and the results summed to approximate a single convolution with the desired dependence on signal level. Accuracy of the result may be traded off with computational load by varying the number of signal quantization levels.

A similar procedure may be used to include the dependence of the slope on both level and frequency. That is, the frequency range may also be divided into subranges, and levels within each subrange are convolved with the level and frequency-specific IIR filters.

Again, the results are summed to approximate a single convolution with the desired dependence on signal level and frequency.

One refinement to the method described above for calculating the dispersion of energy along the basilar membrane is to use a novel spreading function which takes into account the inherent spreading of energy in the frequency domain introduced by the windowing

function used prior to the transform (e.g. Hann Window preceding the Fast Fourier Transform). The shape of this novel spreading function is such that the combined spreading, on the basilar membrane, of the transform window and that of the novel spreading function is equal to the fixed or level and frequency dependent spreading functions described above.

Cognitive Processing

Since the basilar membrane representation produced by the model is expected to represent only supraliminal aspects of the audio signal, this information is the basis for simulating results of listening experiments. However, the perceptual salience of audible basilar degradations can vary depending on a number of contextual factors. Therefore, the reference basilar membrane representation and the basilar degradation vectors are processed in various ways according to reasonable assumptions about human cognitive processing. The result is a number of variables, described below, that together produce a perceptual quality rating.

These are average distortion level, maximum distortion level, average reference level, reference level at maximum distortion, coefficient of variation of distortion, correlation between reference and distortion patterns, and harmonic structure in the distortion.

Other methods also calculate a quality measurement using one or more variables derived from a basilar membrane representation (e.g., [11][12]). These methods use different variables and combinations of variables to produce a quality measurement than are presented here. The variables described below are novel and have not been used previously to measure audio quality.

A value for each variable is computed for each of a discrete number of adjacent frequency ranges. This allows the values for each range to be weighted independently, and also allows interactions among the ranges to be weighted. Three ranges are usually employed - 0 to 1000 Hz, 1000 to 5000 Hz, and 5000 to 18000 Hz. An exception is the measure of harmonic structure of spectrum error that is calculated using the entire audible range of frequencies.

A total of 19 variables result from the seven features listed above when the three pitch regions are taken into account. The variables are mapped to a mean quality rating of that audio sequence as measured in listening tests. Non-linear interactions among the variables are required because the average and maximum errors should be weighted differentially as a function of the coefficient of variation. A multilayer neural network with semi-linear activation functions was applied to allow this possibility.

The feature calculations and the mapping process implemented by the neural network constitute a task-specific model of auditory cognition.

Average Distortion Level

For each analysis frame, the model provides a basilar error vector that describes the extent of degradation over the entire range of auditory frequencies. A positive error represents energy added to the reference signal, while a negative error represents energy

taken away. A single scalar estimate of degradation for the entire sequence of frames could be obtained by integrating the vector elements over time and frequency. However, the perceptibility of distortions is likely modified by the characteristics of the current distortion as well as temporally adjacent distortions. The measured error was modified according to the following criteria.

Perceptual Inertia

A particular distortion is considered inaudible if it is not consistent with the immediate context provided by preceding distortions. This effect will be called perceptual inertia. That is, if the sign of the current error is opposite to the sign of the average error over a short time interval, the error is considered inaudible. The duration of this memory is close to 80 msec, which is the approximate time for the asymptotic integration of loudness of a constant energy stimulus by human listeners – reference [6].

In practice, the energy is accumulated over time, and data from several successive frames determine the state of the memory. At each time step, the window is shifted one frame and each basilar degradation component is summed algebraically over the duration of the window. Clearly, the magnitudes of the window sums depend on the size of the distortions, and whether their signs change within the window. The signs of the sums indicate the state of the memory at that extended instant in time.

The content of the memory is updated with the distortions obtained from processing the current frame. However, the distortion that is output at each time step is the rectified input, modified according to the relation of the input to the signs of the window sums. If the input distortion is positive and the same sign as the window sum, the output is the same as the input. If the sign is different, the corresponding output is set to zero since the input does not continue the trend in the memory at that position. In particular, the output distortion at the ith position, D_i, is assigned a value depending on the sign of the ith window mean, W_i and the ith input distortion, E_i.

$$\text{If (SGN(} E_i \text{) EQ SGN(} W_i \text{) AND } E_i \text{ GT 0.0)} \qquad D_i = E_i$$

$$\text{If (SGN(} E_i \text{) NE SGN(} W_i \text{))} \qquad D_i = 0.0$$

Perceptual Asymmetry

Negative distortions are treated somewhat differently. There are indications in the literature on perception - references [2][4] - that information added to a visual or auditory display is more readily identified than information taken away. Accordingly, this program weighs less heavily the relatively small distortions resulting from spectral energy removed from, rather than added to, the signal being processed. Because it is considered less noticeable, a small negative distortion receives less weight than a positive distortion of the same magnitude. As the magnitude of the error increases, however, the importance of the sign of the error should decrease. The size of the error at which the weight approaches unity was somewhat arbitrarily chosen to be *Pi*, as shown in the following equation.

If (SGN(E_i) EQ SGN(W_i) AND E_i LT 0.0)

$$D_i = |E_i| * \arctan(0.5 * |E_i|)$$

where | | represents the absolute value and * is the scalar multiplication.

Adaptive Threshold for Averaging

The distortion values obtained from the memory could be reduced to a scalar simply by averaging. However, if some pitch positions contain negligible values, the impact of significant adjacent narrow band distortions would be reduced. Such biasing of the average could be prevented by ignoring all values under a fixed threshold, but frames with all distortions under that threshold would then have an average distortion of zero. This also seems like an unsatisfactory bias. Instead, an adaptive threshold has been chosen for ignoring relatively small values. That is, distortions in a particular pitch range are ignored if they are less than one-tenth of the maximum in that range.

The average distortion over time for each pitch range is obtained by summing the mean distortion across successive non-zero frames. A frame is classified as non-zero when the sum of the squares of the most recent 1024 input samples exceeds 8000 (i.e., more than 9 dB per sample on average).

Maximum Distortion Level

The maximum distortion level is obtained independently for each pitch region by finding the frame with the maximum distortion in that range. The maximum value is emphasized for this calculation by defining the adaptive threshold as one-half of the maximum value in the given pitch range instead of one-tenth that is used above to calculate the average distortion.

Average Reference Level

The average reference level over time is obtained by averaging the mean level of the reference signal in each pitch range across successive non-zero frames.

Reference Level at Maximum Distortion

The value of this variable in each pitch region is the reference level that corresponds to the maximum distortion level calculated as described above.

Coefficient of Variation of Distortion

The coefficient of variation is a descriptive statistic that is defined as the ratio of the standard deviation to the mean [10]. The coefficient of variation of the distortion over frames has a relatively large value when a brief, loud distortion occurs in an audio sequence that otherwise has a small average distortion. In this case, the standard deviation is large compared to the mean. Since listeners tend to base their quality judgments on this brief but loud event rather than the overall distortion, the coefficient of

variation may be used to differentially weight the average distortion versus the maximum distortion in the audio sequence. It is calculated independently for each pitch region.

Similarity of reference and distortion spectra

When the peak magnitudes of the distortion coincide in pitch with the peak magnitudes of the reference signal, perceptibility of the distortion may be differentially affected. The correlation between the distortion and reference vectors should reflect this coincidence, and this is found by calculating the cosine of the angle between the vectors for each pitch region as follows:

$$C = \frac{\vec{R} \bullet \vec{E} \bullet \vec{E}}{|\vec{R}| \times |\vec{E}| \times |\vec{E}|}$$

where • is the dot product operator, | | is the magnitude of the enclosed vector and * is again the scalar multiplication.

Harmonic structure in distortion

Listeners may respond to some structure of the error within a frame, as well as to its magnitude. Harmonic structure in the error can result, for example, when the reference signal has strong harmonic structure, and the signal under test includes additional broadband noise. In that case, masking is more likely to be inadequate at frequencies where the level of the reference signal is low between the peaks of the harmonics. The result would be a periodic structure in the noise that corresponds to the structure in the original signal.

The harmonic structure is measured in either of two ways. In the first method, it is described by the location and magnitude of the largest peak in the spectrum of the log energy autocorrelation function. The correlation is calculated as the cosine between two vectors.

In the second method, the periodicity and magnitude of the harmonic structure is inferred from the location of the peak with the largest value in the cepstrum of the error. The relevant parameter is the magnitude of the largest peak. In some cases, it is useful to set the magnitude to zero if the periodicity of the error is significantly different from that of the reference signal. Specifically, if the difference between the two periods is greater than one-quarter of the reference period, the error is assumed to have no harmonic structure related to the original signal.

Additional Useful Variables

Distortion Activity

A distortion that fluctuates a great deal is likely perceived differently than one that is maintained over time. Distortion activity is captured by counting the frequency of reversals of the frame distortion time series over the whole audio sample. A reversal is counted when the change in direction is greater than a critical amount. The reversal

frequency is normalized by dividing by the number of frames in the sequence. For example, a value of .40 indicates that the noise level was more constant over successive frames than a value of .60.

Distortion Clumping

Averaging over time and frequency can cause obvious localized distortions to be under-represented in the final result. To overcome this effect, a parameter is required that is sensitive to the distribution of distortions in the time-frequency plane. A clumpiness parameter is proposed as follows.

$$C = \frac{\sum_{i}^{nPitch} \sum_{j}^{nFrames} d_{i,j} * d_{i,j+1} + d_{i,j} * d_{i+1,j}}{nPitch * nFrames}$$ **Equation I**

The sum of the scalar products of each distortion element $d_{i,j}$ with its neighbour $d_{i,j+1}$ and $d_{i+1,j}$ is accumulated, and this sum is normalised by division by the number of elements *(nPitch x nFrames)*. The result is a value that is sensitive to the degree of aggregation of distortions. The larger the value of the parameter, the more the distortion is localised in frequency and time. Conversely, smaller parameter values are associated with independent errors sprinkled throughout the time-frequency plane.

Auditory Image Shift

Listener ratings may be influenced by changes in binaural relationships that result in apparent movement of the spatial position of the auditory image. In particular, this can occur when phase and level differences between stereo channels change.

The phase distortion is measured by determining the phase for each left and right frequency component using the complex output of the FFT, and measuring the phase lag between the two channels. This phase lag is transformed to an equivalent time lag for each frequency component up to 8 kHz, and the average time lag difference between the reference and coded signals is measured. Similarly, the interaural level distortion is defined as the change in basilar sensation level difference between the left and right channels, also averaged over frequencies up to 8 kHz.

Both phase and level distortions are ignored when the reference basilar sensation level is less than 40 dB relative to a level of one bit.

A related binaural effect is stereo spatiality, or a feeling that the auditory image has depth or three-dimensionality. For example, a single speaking voice typically has little stereo imaging since it is very localised in space. On the other hand, an orchestra or the sound of a rain storm can appear to have considerable depth. The insertion or removal of a sensation of depth is measured by computing the cross-spectrum between the left and right channels of a stereo signal. The cross-spectrum is simply the cross-correlation

between the complex spectra. A change in stereo imaging is indicated by a change in the cross-spectrum, especially for frequencies ranging from approximately 3 to 10 kHz.

Pre-echo Detection

Pre-echo is a coding artifact that appears immediately preceding a sudden increase, or attack, in the audio signal. An attack detector was developed to spot sudden changes in the average signal level. The main problem was how to measure the average signal energy. If the averaging is done over too large a window, the temporal resolution is too large to determine the rate of increase in the signal strength. On the other hand short-term averages are rather unstable, resulting in many false alarms. The ideal attack detector would use a non-linear filter that responds quickly to rising signal level but decays slowly during decreasing signal strength.

A nonlinear filter with satisfactory properties is the simple maximum operator. Given a window of several thousand samples, it returns the maximum of the absolute (rectified) signal level. It has instant response when a large signal enters the window. In order to avoid reporting transients separated by less than 100 milliseconds, a window corresponding to 1/10 of a second (4800 samples at a 48 kHz sampling rate) was chosen. To avoid very low frequency effects, the incoming signal was differentiated and rectified.

Computing the maximum of 4800 samples for each sample shift of the window can impose a large computational load if it is not done efficiently. To minimize this problem the incoming signal is decimated by a factor of four. The window is also partitioned into 20 smaller windows and the maximum is computed both for each partition and for all the partitions. A new sample only effects one partition, so the other partitions do not need to be recomputed.

Calibration of Model Predictors

The mean quality ratings obtained from human listening experiments is predicted by a weighted non-linear combination of 6 groups of variables consisting of

- average basilar degradation for three equal pitch ranges
- average reference signal level for each pitch range
- maximum basilar degradation for each pitch range
- the reference signal level at maximum degradation
- the coefficient of variation of the basilar degradation for each pitch range
- average correlation between the distortion and the reference signal
- average harmonic structure in the distortion

The prediction algorithm was optimised using a multilayer neural network to derive the appropriate weightings of the input variables. This method permits non-linear interactions among the variables which is required to differentially weight the average distortion and the maximum distortion as a function of the coefficient of variation.

The system relating the above variables to human quality ratings was calibrated using data from eight different listening tests that used the same basic methodology. These experiments are known in the ITU-R Task Group 10/4 as MPEG90, MPEG91, ITU92CO, ITU92DI, ITU93, MPEG95, EIA95, and DB2. Generalisation testing was performed using data from the DB3 and CRC97 listening tests.

Source Codes

A set of source code listings for software used in the present invention is attached as Appendix A.

References

[1] M. Florentine and S. Buus. An excitation-pattern model for intensity discrimination. *J. Acoust. Soc. Am.*, 70:1646-1654, 1981.

[2] E. Hearst. Psychology and nothing. *American Scientist*, 79:432-443, 1979.

[3] E. Terhardt, G. Stoll, M. Sweeman. Algorithm for extraction of pitch and pitch salience from complex tonal signals. *J. Acoust. Soc. Am.* 71(3):678-688, 1982.

[4] M. Treisman. Features and objects in visual processing. *Scientific American*, 255[5]:114-124, 1986.

[5] E. Zwicker and E. Terhardt. Analytical expressions for critical-band rate and critical bandwidth as a fuction of frequency. *J. Acoust. Soc. Am.* 68(5): 1523-1525, 1980.

[6] Treurniet, W.C. Simulation of individual listeners with an auditory model. *Proceedings of the Audio Engineering Society*, Copenhagen, Denmark, Reprint Number 4154, 1996.

[7] B. Paillard, P. Mabilleau, S. Morisette, and J. Soumagne. Perceval: Perceptual evaluation of the quality of audio signals, *J. Audio Eng. Soc.*, Vol. 40, pages 21-31, 1992.

[8] J.G. Beerends and J.A. Stemerdink, A perceptual audio quality measure based on a psychoacoustic sound representation, *J. Audio Eng. Soc.*, Vol. 40, pages 963-978, December 1992.

[9] C. Colomes, M. Lever, J.B. Rault, and Y.F. Dehery, A perceptual model applied to audio bit-rate reduction, *J.Audio Eng. Soc.*, Vol. 43, pages 233-240, April 1995.

[10] K. Brandenburg and T. Sporer. 'NMR' and 'Masking Flag': Evaluation of quality using perceptual criteria, *11th International AES Conference on Audio Test and Measurement*, Portland, 1992, pp169-179.

[11] T. Thiede and E. Kabot, A New Perceptual Quality Measure for Bit Rate Reduced Audio *Proceedings of the Audio Engineering Society*, Copenhagen, Denmark, Reprint Number 4280, 1996.

[12] J.G. Beerends, Measuring the quality of speech and music codecs, an integrated psychoacoustic approach. *Proceedings of the Audio Engineering Society*, Copenhagen, Denmark, Reprint Number 4154, 1996.

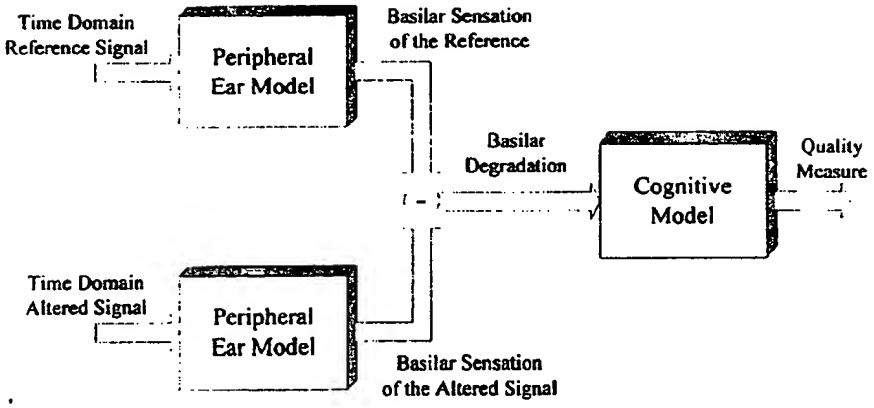


Figure 1

Scott & Aylen

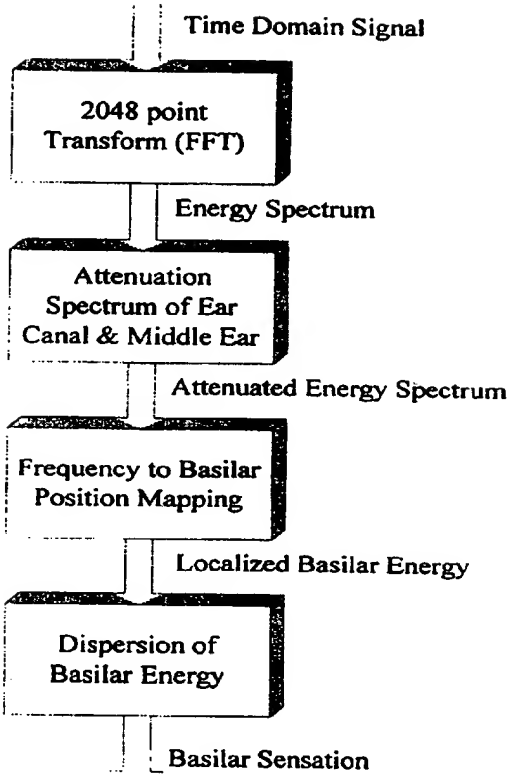


Figure 2.

Scott & Aylen